International Journal of Engineering Research & Management Technology ISSN: 2348-4039

Email: editor@ijermt.org November- 2014 Volume 1, Issue-6 www.ijermt.org

Layered Approach Using Conditional Random Fields for Intrusion Detection

Ranjeet Singh, Chiranjit Dutta
Faculty of Information Technology
SRM University
NCR Campus, Ghaziabad

ABSTRACT

Intrusion detection faces a number of challenges. An intrusion detection system must reliably detect malicious activities in a network and must perform efficiently to cope with the large amount of network traffic. In this paper, we address these two issues of Accuracy and Efficiency using Conditional Random Fields and Layered Approach. We demonstrate that high attack detection accuracy can be achieved by using Conditional Random Fields and high efficiency by implementing the Layered Approach. Our proposed system based on Layered Conditional Random Fields outperforms other well-known methods such as the decision trees and the naive Bayes. The improvement in attack detection accuracy is very high, particularly, for the U2R attacks (34.8 percent improvement) and the R2L attacks (34.5 percent improvement). Statistical Tests also demonstrate higher confidence in detection accuracy for our method. Finally, we show that our system is robust and is able to handle noisy data without compromising performance.

Keywords - Intrusion detection, Layered Approach, Conditional Random Fields, Network Security.

1. INTRODUCTION

With the increase in the use of internet the concerns of making the internet more secure were also emerged among the technocrats and users. Because of this concern, many intrusion detection techniques came into existence. Though, IDS is considered to be immature and it does not provide a complete defense, but we believe that it can play a significant role in overall security architecture. As in battle field a warning can play a major role, similarly a warning can provide alert to the user about any skeptical attack on the system, hence, this warning indication that the system is under attack, even if the system is not assailable to specific attack, can help users to revamp their installation's defensive posture and can increase resistance to attack. Intrusion detection systems mainly base their decisions either on Signal (signature-based detection) or Noise (anomaly-based detection). IDS can also be classified on the phenomenology that they sense. Network-based system can simultaneously monitor numerous hosts; they can suffer from performance problems, especially with increasing network speeds. Another is host-based system that can monitor specific applications in ways that would be difficult or impossible in a network-based system[13]. While there is an existence of Hybrid System, this system is the combination of both signature-based and the anomaly-based systems. Hybrid Systems system can be very efficient when subjected to classification methods and can also be used to label unseen (new instances) as they assign one of the known classes to every test instance. This is because during training the system learns features from all the classes[11]. In this paper, we are trying to make a result oriented comparison among different combinational models, which are created by us with layered approach and results are used to find best combinational method for all types of attacks.

2. RELATED WORK

This detection approach was employed to detect attack categories in the NSL-KDD dataset. The technique has achieved the detection rate of 97.48% for DOS, 95.23% for Probe, 99.49% for U2R and 96.48% of R2L respectively. This statics shows that our approach is very much accurate for every type of attack. In 1997[5], Richard Maclin and David Optiz, presented "An empirical evaluation of boosting and bagging" in which they have shown that when these meta-algorithms are used they produce a larger gain in accuracy. This encouraged many researchers and then in 2008[8] Weiming Hu and Wei Hu presented "An intrusion detection system using ad boost meta-algorithm" which also shows a significant or competitive performance with IDS systems. In 2010[2], Kapil Kumar Gupta, Baikunth Nath and Ramamohanaroa Kotagiri presented "A frame work using a layered approach for intrusion detection". They have addressed two main issues of ID i.e. accuracy and efficiency by using conditional random fields and layered approach. They have shown that layered CRFs have very high attack detection rate 98.6% for probe and 97.40% for DOS. However, they were outperformed by a significant percent for the R2L and U2R attacks. Where, our approach performs fantastically. We are also influenced by the work of [9], [12] [5] and many more authors. Literature survey has shown us that in all particular purposes most of the researchers have applied a single algorithm to address all the four attack categories. This has motivated us and helps us to draw an assumption, that the combination of different algorithms would perform different predictions on different attack categories, and may yield a good performance and high prediction comparatively.

3. INTRUSION DETECTION

Intrusion detection as defined by the System Administrators, Audit, Networking and Security (SANs) Institute is the art of detecting inappropriate, inaccurate or anomalous activity. We use intrusion detection systems to protect our network from attacks and abuses, we also use it to detect the violation in security and attacks on network, to document them and to get detailed information about intrusions that occurred [13]. There are following approaches for IDS:

a) Signature-based approach: Design to detect the known attacks. It is very effective for detecting the attacks without generating an overwhelming number of false alarms; it can quickly and reliably diagnose the use of a specific attack tool. But it has a loophole, that it can only detect the attacks which are described in its database.

b) Classification-based approach: This approach uses normal and abnormal datasets of user behavior and uses data mining techniques to train the IDS system. This creates more accurate classification models for IDS as compared to signature-based approaches and thus they are more powerful in detecting known attacks. But still they are not capable of detecting unknown attacks.

c) Anomaly-based approach: The basic assumption of anomaly detection approach is that, attacks are different from normal activities and thus they can be detected by IDS systems that identify these differences. This detection approach can detect unknown attacks also, but still it has a loophole, this approach generates a large number of false alarms due to unpredictable behaviors of users and networks. Data mining approaches are relatively new technique for intrusion detection. There are a wide variety of data mining algorithms drawn from the fields of statics, pattern recognition, machine learning and database.

4. CONDITIONAL RANDOM FIELDS FOR INTRUSION DETECTION

Conditional models are probabilistic systems that are used to model the conditional distribution over a set of random variables. Such models have been extensively used in the natural language processing tasks. Conditional models offer a better framework as they do not make any unwarranted assumptions on the observations and can be used to model rich overlapping features among the visible observations. The advantage of CRFs is that they are undirected and are, thus, free from the Label Bias and the Observation Bias. The simplest conditional classifier is the Maxent classifier based upon maximum entropy classification, which estimates the conditional distribution of every class given the observations. The training data is used to constrain this conditional distribution while ensuring maximum entropy and hence maximum uniformity.

5. LAYERED APPROACH

Layered-based intrusion detection system gets its motivation from Airport security model, where a number of security checks are performed one after the other in sequence. Similar to this model, the layered intrusion detection system represents a sequential layered approach and is based on ensuring clandestinely, credibility and availability

of data or is significant services over a network. The goal of using a layered model is to reduce computation and the overall time required to detect anomalous events. The time required to detect an intrusive event is significant and can be reduced by eliminating the communication overhead among different layers. This can be achieved by making the layers autonomous and self-sufficient to block an attack without the need of a central decision maker. Every layer in layered intrusion detection system framework is trained separately and then deployed sequentially. We define four layers that correspond to the four attack groups mentioned in the dataset. They are probe layer, DOS layer, U2R layer and R2L layer. Each layer is then separately trained with a small set of relevant features.
Feature selection or reduction is important for layered approach and discussed in next section. In order to make the layers independent, some features may be present in more than one layer. The layers essentially act as filters that block any anomalous connection, thereby eliminating the need of further processing at subsequent layers enabling quick response to intrusion. The effect of such a sequence of layers is that the anomalous events are identified and blocked as soon as they are detected [2].

6. INTRUSION DATA

During attack, an attacker sets up a connection between a source IP address to a target IP address and sends data to attack the target. The simulated attacks fall in one of the following categories:

- DOS (Denial of service) in this type of attacks an attacker makes some computing or memory resources too busy to handle legitimate requests.
- Probing attack, in this attacker scans a network of computers to gather information and then uses it to exploit the system.
- U2R (User to root attack), in this an attacker starts out with access to a normal user account on the system and is able to exploit vulnerability to gain root access to the system.
- R2L (Remote to local attack), in this an attacker who does not have an account on a remote machine sends packet to that machine over a network and exploits some vulnerability to gain local access.

7. RESULTS AND ANALYSIS

In Our experiment we use NSL-KDD datasets. Due to some inherent problems of KDD¡¦99 dataset, NSL-KDD comes into existence. The number of records in the NSL-KDD train and test sets are reasonable. This advantage makes it affordable to run the experiments on the complete set without the need of randomly selecting a small portion. The training datasets of NSL-KDD are similar to KDD¡¦99 and consist of approximately 4,900,000 single connection vectors each of them contains 41 features and is labeled as normal or attack type with exactly one specific attack type.

NSL-KDD datasets have following advantages over the original KDD datasets:

- Train sets are free from redundant records, which results into unbiased classifier.
- Duplicate records are not present in the test set, which results into unbiased performance of learner.
- The classification rates of distinct machine learning methods vary in a wider range, which makes NSL-KD datasets more efficient to have an accurate evaluating different learning techniques.
- We compare our work with other well-known methods based on the anomaly intrusion detection principle. The anomaly-based systems primarily detect deviations from the learnt normal data by using statistical methods, machine learning, or data mining approaches. Standard techniques such as the decision trees and naive Bayes are known to perform well.
- The most impressive part of the Layered CRFs is the margin of improvement as compared with other methods. Layered CRFs have very high attack detection of 98.6 percent for Probes (5.8 percent improvement) and 97.40 percent detection for DoS. They outperform by a significant percentage for the R2L (34.5 percent improvement) and the U2R (34.8 percent improvement) attacks.

8. CONCLUSION

In this paper, we have addressed aspects of intrusion detection system and made it more robust and efficient; one aspect is accuracy and other is performance. Our experimental results have shown that, "A layered approach for intrusion detection using meta-modeling with classification techniques" is very effective in improving the attack detection rate. The area for future research include, finding the robustness of our system with noisy dataset. As well as trying to find a more effective feature reduction approach.

REFERENCE

1. Shun-ichi Amari and Si Wu, "Improving support vector machine classifiers by modifying kernel function", RIKEN Brain Science Institute Japan.
2. Kapil Kumar Gupta, Baikunth Nath and Ramamohanaroo kotagiri, "A layered approach using conditional random fields for intrusion detection", IEEE Tranc. on Dependence and secure computing, Vol.7,2010
3. G.MeeraGandhi, Kumaravel Appavoo and S.K Srivasta, "Effective network intrusion detection using classifiers decision trees and decision rules",Int. J. Advanced network and application, Vol2, 2010
4. Sandy Peddabachigari, Ajit Abraham and Johmson Thomas, "Intrusion detection system using decision trees and SVM", Oklahoma state university USA.
6. Huy Anh Nguyen and Deokjai choi, "Application of data mining to network intrusion detection", Korea.
7. Weiming Hu, Wei Hu and Steve Maybank, "Adaboost based algorithm for network intrusion detection", Tranc. On system man and cybernetics, 2008.
8. Shilpa Lakhina, Sini Joseph and Bhupendra Verma, "Feature reduction using PCA for effective Anomaly-based intrusion detection on NSL-KDD", Int. J. of engineering science and technology, 2010
9. nehal A.Mulay, P.R Devale and G.V Garje, "Intrusion detection using SVM and decision tree", Int. J. of computer application, 2010
10. J.Vishumathi and K.L Shunmuganathan, "A computational intelligence for evaluation of intrusion detection system ", Indian J. of science and technology, Jan 2011
11. detection system", Int. J. of soft computing and engineering, May 2011
12. Xunyi Ren, Ruchuan Wang and Hejunzhou, "intrusion detection system method using protocol classification and Rough set based SVM", www.ccsenet.org/journal.html,2009
13. Peyman Kabiri and Ali A. Ghorbani, "Research on ID and Response: A survey ", Int. J. of network security, 2005
14. "Bagging and boosting", Srihari@cedar.buffalo.edu
15. Tich Phuoc Tran, Longbing cao, Dat Tran and Cu ong Duc Nguyen, "Novel ID using probabilistic Neural network and adaptive boosting", Int. J. of CS and information security, 2009
16. NSL-KDD.html
17. Lindsay I Smith A tutorial on Principal Components Analysis February 26,2002.
18. A. B. M. S. Ali, A. Abraham. An Empirical Comparison of Kernel Selection for Support Vector Machines. 2nd International Conference on Hybrid Intelligent Systems: Design, Management and Applications,The Netherlands, 2002
19. Robi Polikar, "Ensemble based systems in decision making", IEEE circuit and system magazine, 2006
20. Jaiwei Han and Micheline Kamber, Elseiver book
21. P.Garcia- Teodoro, J.Diaz- Verdejo, "Anomalyy network intrusion detection: Techniques, systems and challenges", www.elsevier.com ,2009